

BC#6

PCT09

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/830,647

DATE: 08/30/2001

TIME: 07:37:38

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

ENTERED

```

3 <110> APPLICANT: ARAI, Kenichi
4   MASAI, Hisao
6 <120> TITLE OF INVENTION: Human H37 Protein and cDNA Encoding The Protein
8 <130> FILE REFERENCE: 2001-0531A/WMC/00653
10 <140> CURRENT APPLICATION NUMBER: 09/830,647
11 <141> CURRENT FILING DATE: 2001-04-30
13 <150> PRIOR APPLICATION NUMBER: JP No. 10-311408
14 <151> PRIOR FILING DATE: 1998-10-30
16 <160> NUMBER OF SEQ ID NOS: 4
18 <210> SEQ ID NO: 1
19 <211> LENGTH: 674
20 <212> TYPE: PRT
21 <213> ORGANISM: Homo sapiens
23 <400> SEQUENCE: 1
24 Met Asn Ser Gly Ala Met Arg Ile His Ser Lys Gly His Phe Gln Gly
25   1           5           10           15
26 Gly Ile Gln Val Lys Asn Glu Lys Asn Arg Pro Ser Leu Lys Ser Leu
27   20           25           30
28 Lys Thr Asp Asn Arg Pro Glu Lys Ser Lys Cys Lys Pro Leu Trp Gly
29   35           40           45
30 Lys Val Phe Tyr Leu Asp Leu Pro Ser Val Thr Ile Ser Glu Lys Leu
31   50           55           60
32 Gln Lys Asp Ile Lys Asp Leu Gly Gly Arg Val Glu Glu Phe Leu Ser
33   65           70           75           80
34 Lys Asp Ile Ser Tyr Leu Ile Ser Asn Lys Lys Glu Ala Lys Phe Ala
35   85           90           95
36 Gln Thr Leu Gly Arg Ile Ser Pro Val Pro Ser Pro Glu Ser Ala Tyr
37   100          105          110
38 Thr Ala Glu Thr Thr Ser Pro His Pro Ser His Asp Gly Ser Ser Phe
39   115          120          125
40 Lys Ser Pro Asp Thr Val Cys Leu Ser Arg Gly Lys Leu Leu Val Glu
41   130          135          140
42 Lys Ala Ile Lys Asp His Asp Phe Ile Pro Ser Asn Ser Ile Leu Ser
43   145          150          155          160
44 Asn Ala Leu Ser Trp Gly Val Lys Ile Leu His Ile Asp Asp Ile Arg
45   165          170          175
46 Tyr Tyr Ile Glu Gln Lys Lys Lys Glu Leu Tyr Leu Leu Lys Lys Ser
47   180          185          190
48 Ser Thr Ser Val Arg Asp Gly Gly Lys Arg Val Gly Ser Gly Ala Gln
49   195          200          205
50 Lys Thr Arg Thr Gly Arg Leu Lys Lys Pro Phe Val Lys Val Glu Asp
51   210          215          220
52 Met Ser Gln Leu Tyr Arg Pro Phe Tyr Leu Gln Leu Thr Asn Met Pro
53   225          230          235          240
54 Phe Ile Asn Tyr Ser Ile Gln Lys Pro Cys Ser Pro Phe Asp Val Asp
55   245          250          255
56 Lys Pro Ser Ser Met Gln Lys Gln Thr Gln Val Lys Leu Arg Ile Gln

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/830,647

DATE: 08/30/2001

TIME: 07:37:38

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

```

57          260          265          270
58 Thr Asp Gly Asp Lys Tyr Gly Gly Thr Ser Ile Gln Leu Gln Leu Lys
59          275          280          285
60 Glu Lys Lys Lys Lys Gly Tyr Cys Glu Cys Cys Leu Gln Lys Tyr Glu
61          290          295          300
62 Asp Leu Glu Thr His Leu Leu Ser Glu Gln His Arg Asn Phe Ala Gln
63          305          310          315          320
64 Ser Asn Gln Tyr Gln Val Val Asp Asp Ile Val Ser Lys Leu Val Phe
65          325          330          335
66 Asp Phe Val Glu Tyr Glu Lys Asp Thr Pro Lys Lys Lys Arg Ile Lys
67          340          345          350
68 Tyr Ser Val Gly Ser Leu Ser Pro Val Ser Ala Ser Val Leu Lys Lys
69          355          360          365
70 Thr Glu Gln Lys Glu Lys Val Glu Leu Gln His Ile Ser Gln Lys Asp
71          370          375          380
72 Cys Gln Glu Asp Asp Thr Val Lys Glu Gln Asn Phe Leu Tyr Lys
73          385          390          395          400
74 Glu Thr Gln Glu Thr Glu Lys Lys Leu Leu Phe Ile Ser Glu Pro Ile
75          405          410          415
76 Pro His Pro Ser Asn Glu Leu Arg Gly Leu Asn Glu Lys Met Ser Asn
77          420          425          430
78 Lys Cys Ser Met Leu Ser Thr Ala Glu Asp Asp Ile Arg Gln Asn Phe
79          435          440          445
80 Thr Gln Leu Pro Leu His Lys Asn Lys Gln Glu Cys Ile Leu Asp Ile
81          450          455          460
82 Ser Glu His Thr Leu Ser Glu Asn Asp Leu Glu Glu Leu Arg Val Asp
83          465          470          475          480
84 His Tyr Lys Cys Asn Ile Gln Ala Ser Val His Val Ser Asp Phe Ser
85          485          490          495
86 Thr Asp Asn Ser Gly Ser Gln Pro Lys Gln Lys Ser Asp Thr Val Leu
87          500          505          510
88 Phe Pro Ala Lys Asp Leu Lys Glu Lys Asp Leu His Ser Ile Phe Thr
89          515          520          525
90 His Asp Ser Gly Leu Ile Thr Ile Asn Ser Ser Gln Glu His Leu Thr
91          530          535          540
92 Val Gln Ala Lys Ala Pro Phe His Thr Pro Pro Glu Glu Pro Asn Glu
93          545          550          555          560
94 Cys Asp Phe Lys Asn Met Asp Ser Leu Pro Ser Gly Lys Ile His Arg
95          565          570          575
96 Lys Val Lys Ile Ile Leu Gly Arg Asn Arg Lys Glu Asn Leu Glu Pro
97          580          585          590
98 Asn Ala Glu Phe Asp Lys Arg Thr Glu Phe Ile Thr Gln Glu Glu Asn
99          595          600          605
100 Arg Ile Cys Ser Ser Pro Val Gln Ser Leu Leu Asp Leu Phe Gln Thr
101          610          615          620
102 Ser Glu Glu Lys Ser Glu Phe Leu Gly Phe Thr Ser Tyr Thr Glu Lys
103          625          630          635          640
104 Ser Gly Ile Cys Asn Val Leu Asp Ile Trp Glu Glu Glu Asn Ser Asp
105          645          650          655

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/830,647

DATE: 08/30/2001

TIME: 07:37:38

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

106 Asn Leu Leu Thr Ala Phe Phe Ser Ser Pro Ser Thr Ser Thr Phe Thr
 107 660 665 670

108 Gly Phe

109 674

111 <210> SEQ ID NO: 2

112 <211> LENGTH: 234

113 <212> TYPE: PRT

114 <213> ORGANISM: Homo sapiens

116 <400> SEQUENCE: 2

117 Met Asn Ser Gly Ala Met Arg Ile His Ser Lys Gly His Phe Gln Gly

118 1 5 10 15

119 Gly Ile Gln Val Lys Asn Glu Lys Asn Arg Pro Ser Leu Lys Ser Leu

120 20 25 30

121 Lys Thr Asp Asn Arg Pro Glu Lys Ser Lys Cys Lys Pro Leu Trp Gly

122 35 40 45

123 Lys Val Phe Tyr Leu Asp Leu Pro Ser Val Thr Ile Ser Glu Lys Leu

124 50 55 60

125 Gln Lys Asp Ile Lys Asp Leu Gly Gly Arg Val Glu Glu Phe Leu Ser

126 65 70 75 80

127 Lys Asp Ile Ser Tyr Leu Ile Ser Asn Lys Lys Glu Ala Lys Phe Ala

128 85 90 95

129 Gln Thr Leu Gly Arg Ile Ser Pro Val Pro Ser Pro Glu Ser Ala Tyr

130 100 105 110

131 Thr Ala Glu Thr Thr Ser Pro His Pro Ser His Asp Gly Ser Ser Phe

132 115 120 125

133 Lys Ser Pro Asp Thr Val Cys Leu Ser Arg Gly Lys Leu Leu Val Glu

134 130 135 140

135 Lys Ala Ile Lys Asp His Asp Phe Ile Pro Ser Asn Ser Ile Leu Ser

136 145 150 155 160

137 Asn Ala Leu Ser Trp Gly Val Lys Ile Leu His Ile Asp Asp Ile Arg

138 165 170 175

139 Tyr Tyr Ile Glu Lys Lys Lys Glu Leu Tyr Leu Leu Lys Lys Ser

140 180 185 190

141 Ser Thr Ser Val Arg Asp Gly Gly Lys Arg Val Gly Ser Gly Ala Gln

142 195 200 205

143 Lys Thr Arg Thr Gly Arg Leu Lys Lys Pro Phe Val Lys Val Glu Asp

144 210 215 220

145 Met Ser Gln Ser Pro Ala Val His Leu Met

146 225 230 234

148 <210> SEQ ID NO: 3

149 <211> LENGTH: 2780

150 <212> TYPE: DNA

151 <213> ORGANISM: Homo sapiens

153 <400> SEQUENCE: 3

154 aattcggcac gagctctctg aggctgcgcc aagacctgaa gcggcgacc gagagcccg 60

155 gtctgagact gagagagcaa cggaatggag gcggggtaga ggcgaaaca caacctgcag 120

156 ggccagagcg aggcgcgaga aggacggcgg cgtgaggggg cggggcgcgc agcgcgagaa 180

157 ggcagggcag aggggcgagc gcgagggcgg gcacggcgcg tggcgtgaga cggggcgggg 240

158 cgcgcgtatc ggcgcgcggc ccgcgtgacg cgttttcaaa tcttcaaccg ccgcagccca 300

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/830,647

DATE: 08/30/2001

TIME: 07:37:38

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

```

159 ctcgtttgtg ctttgcgcct tcctcctccg cgccttggag ccggatccgg ccccggaac 360
160 ccgacctgca gacgcggtac ctctactgcg tagaggccgt agctggcgga aggagagagg 420
161 cggccgtcct gtcaacaggc cgggggaagc cgtgctttcg cggtgcccg gtgcgacact 480
162 ttctccggac ccagcatgta ggtgccgggc gactgccatg aactccggag ccatgaggat 540
163 ccacagtaaa ggacatttcc aggggtgaat ccaagtcaaa aatgaaaaaa acagaccatc 600
164 tctgaaatct ctgaaaactg ataacaggcc agaaaaatcc aaatgtaagc cactttgggg 660
165 aaaagtattt taccttgact taccttctgt caccatatct gaaaaacttc aaaaggacat 720
166 taaggatctg ggagggcgag ttgaagaatt tctcagcaaa gatatcagtt atcttatttc 780
167 aaataagaag gaagctaaat ttgcacaaac cttgggtcga atttctctctg taccaagtcc 840
168 agaattctgca tatactgcag aaaccacttc acctcatccc agccatgatg gaagttcatt 900
169 taagtaccca gacacagtgt gtttaagcag agggaaatta ttagttgaaa aagctatcaa 960
170 ggaccatgat tttattcctt caaatagtat attatcaaat gccttgtcat ggggagtaaa 1020
171 aattcttcat attgatgaca ttagatacta cattgaacaa aagaaaaaag agttgtattt 1080
172 actcaagaaa tcaagtactt cagtaagaga tgggggcaaa agagttggta gtggtgcaca 1140
173 aaaaacaaga acaggaagac tcaaaaagcc ttttgtaaag gtggaagata tgagccaact 1200
174 ttataggcca ttttatcttc agctgaccaa tatgcctttt ataaattatt ctattcagaa 1260
175 gccctgcagt ccaattgatg tagacaagcc atctagtatg caaaagcaaa ctcaggttaa 1320
176 actaagaatc caaacagatg gcgataagta tgggtggaacc tcaattcaac tccagttgaa 1380
177 agagaagaag aaaaaaggat attgtgaatg ttgcttgcag aaatatgaag atctagaaac 1440
178 tcaccttcta agtgagcaac acagaaactt tgcacagagt aaccagtatc aagttgttga 1500
179 tgatattgta totaagttag tttttgactt tgtggaatat gaaaaggaca cacctaaaaa 1560
180 gaaaagaata aaatacagtg ttggatccct ttctctggtt tctgcaagtg tcctgaaaaa 1620
181 gactgaacaa aaggaaaaag tgggaattgca acatatctct cagaaagatt gccaggaaga 1680
182 tgatacaaca gtgaaggagc agaatttctt gtataaagag acccaggaaa ctgaaaaaaa 1740
183 gctcctgttt atttcagagc ccattcccca cccttcaaat gaattgagag ggcttaatga 1800
184 gaaaatgagt aataaatggt ccattgttaag tacagctgaa gatgacataa gacagaattt 1860
185 tacacagcta cctctacata aaaacaaaca ggaatgcatt cttgacattt ccgaacacac 1920
186 attaagtga aatgacttag aagaactaag ggtagatcac tataaatgta acatacaggc 1980
187 atctgtacat gtttctgatt tcagtacaga taatagtga tctcaaccaa aacagaagtc 2040
188 agatactgtg ctttttccag caaaggatct caaggaaaag gaccttcatt caatatttac 2100
189 tcatgattct ggtctgataa caataaacag ttcacaagag cacctaactg ttcaggcaaa 2160
190 ggtccatttc catactcctc ctgaggaacc caatgaatgt gacttcaaga atatggatag 2220
191 tttaccttct ggtaaaatac atcgaaaagt gaaaataata ttaggacgaa atagaaaaga 2280
192 aaatctggaa ccaaatgctg aatttgataa aagaactgaa tttattacac aagaagaaaa 2340
193 cagaatttgt agttcaccgg tacagtcttt actagacttg tttcagacta gtgaagagaa 2400
194 atcagaatth ttgggtttca caagctacac agaaaagagt ggtatatgca atgttttaga 2460
195 tttttgggaa gaggaaaatt cagataatct gttaacagcg tttttctcgt ccccttcaac 2520
196 ttctacattt actggctttt agaattttaa aaatgcatac ttttcagaag tgataaggat 2580
197 catattcttg aaatttttat aaatatgtat ggaaattctt aggatttttt taccagcttt 2640
198 gtttacagac ccaaatgtaa atattaaaaa taaatatttg caattttcta cagaattgaa 2700
199 tacctgttaa agaaaaatta cagaataaac ttgtgactgg tcttgtttta cattaaaaaa 2760
200 aaaaaaaaaa aaaactcgag 2780
202 <210> SEQ ID NO: 4
203 <211> LENGTH: 2719
204 <212> TYPE: DNA
205 <213> ORGANISM: Homo sapiens
207 <400> SEQUENCE: 4
208 aattcggcac gagctctctg aggctgcgcc aagacctgaa gcggcggacc gagagcccgg 60
209 gtctgagact gagagagcaa cggaatggag gcggggtaga ggcggaaaca caacctgcag 120

```

RAW SEQUENCE LISTING

DATE: 08/30/2001

PATENT APPLICATION: US/09/830,647

TIME: 07:37:38

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

```

210 ggccagagcg aggcgcgaga aggcgcgcgg cgtgaggggg cgggggcgcgc agcgcgagaa 180
211 ggcagggcacg agggggcgagc gcgagggcggg gcacggcgcg tggcgtgaga cggggcgggg 240
212 cgcgcgtatc ggcgcgcgcgg ccgcgtgacg cgttttcaaa tcttcaaccg ccgcagccca 300
213 ctcgtttggtg ctttgccctt tctctctccg cgccttgag cgggatccgg ccccggaac 360
214 ccgacctgca gacgcggtac ctctactgcg tagaggccgt agctggcggg aggagagagg 420
215 cggccgtcct gtcaacaggc cgggggaagc cgtgctttcg cggctgcccg gtgcgacact 480
216 ttctccggac ccagcatgta ggtgcggggc gactgccatg aaactccggg ccatgaggat 540
217 ccacagtaaa ggacatttcc aggggtggaat ccaagtcaaa aatgaaaaaa acagaccatc 600
218 tctgaaatct ctgaaaactg ataacaggcc agaaaaatcc aaatgtaagc cactttgggg 660
219 aaaagtattt taccttgact taccttctgt caccatatct gaaaaacttc aaaaggacat 720
220 taaggatctg ggagggcgag ttgaagaatt tctcagcaa gatatcagtt atcttatttc 780
221 aaataagaag gaagctaaat ttgcacaaac cttgggtcga atttctcctg taccaagtcc 840
222 agaattctgca tatactgcag aaaccaacttc acctcatccc agccatgatg gaagttcatt 900
223 taagtcacca gacacagtgt gtttaagcag aggaaaatta ttagttgaaa agctatcaa 960
224 ggaccatgat ttatttcctt caaatagtat attatcaaat gccttgatcat ggggagtaaa 1020
225 aattcttcat attgatgaca ttagatacta cattgaacaa aagaaaaaag agttgtattt 1080
226 actcaagaaa tcaagtactt cagtaagaga tgggggcaaa agagttggta gtggtgcaca 1140
227 aaaaacaaga acaggaagac tcaaaaagcc ttttgtaaag gtggaagata tgagccaaag 1200
228 cctgcagtc catttgatgt agacaagcca tctagtatgc aaaagcaaac tcaggttaaa 1260
229 ctaagaatcc aaacagatgg cgataagtat ggtggaacct caattcaact ccagttgaaa 1320
230 gagaagaaga aaaaaggata ttgtgaatgt tgcttgcaaa aatatgaaga tctagaaact 1380
231 cactttctaa gtgagcaaca cagaaacttt gcacagagta accagtatca agttgttgat 1440
232 gatattgtat ctaagttagt ttttgacttt gtggaatatg aaaaggacac acctaaaaag 1500
233 aaaagaataa aatacagtgt tggatccctt tctcctgttt ctgcaagtgt cctgaaaaag 1560
234 actgaacaaa aggaaaaagt ggaattgcaa catatttctc agaaagattg ccaggaagat 1620
235 gatacaacag tgaaggagca gaatttctcg tataaagaga cccaggaaac tgaaaaaaag 1680
236 ctctgtttta ttccagagcc catccccac ctttcaaagt aattgagagg gcttaatgag 1740
237 aaaatgagta ataatgttc catgttaagt acagctgaag atgacataag acagaatttt 1800
238 acacagctac ctctacataa aaacaaacag gaatgcattc ttgacatttc cgaacacaca 1860
239 ttaagtgaag atgacttaga agaactaagg gtagatcact ataaatgtaa catacaggca 1920
240 tctgtacatg tttctgattt cagtacagat aatagtggat ctcaaccaa acagaagtca 1980
241 gatactgtgc tttttccagc aaaggatctc aaggaaaagg accttcattc aatatttact 2040
242 catgattctg gtctgataac aataaacagt tcacaagagc acctaaactgt tcaggcaaag 2100
243 gctccattcc atactcctcc tgaggaaacc aatgaatgtg acttcaagaa tatggatagt 2160
244 ttaccttctg gtaaaatata tcgaaaagtg aaaataatat taggacgaaa tagaaaagaa 2220
245 aatctggaac caaatgctga atttgataaa agaactgaat ttattacaca agaagaaaac 2280
246 agaatttgta gttcaccggg acagctttta ctgacttgt ttcagactag tgaagagaaa 2340
247 tcagaatttt tgggtttcac aagctacaca gaaaagagt gtatatgcaa tgttttagat 2400
248 atttgggaag aggaaaattc agataatctg ttaacagcgt ttttctcgtc ccttcaact 2460
249 totacattta ctggctttta gaatttaaaa aatgcatact tttcagaagt gataaggatc 2520
250 atattcttga aatttttata aatatgtatg gaaattctta ggattttttt accagctttg 2580
251 ttacagacc caaatgtaaa tattaataat aaatatttgc aattttctac agaattgaat 2640
252 acctgttaaa gaaaaattac agaataaact tgtgactggg cttgttttac attaaaaaaa 2700
253 aaaaaaaaaa aaactcgag                                     2719

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/830,647

DATE: 08/30/2001

TIME: 07:37:39

Input Set : A:\ARAI SEQUENCE LISTING.txt

Output Set: N:\CRF3\08302001\I830647.raw

L:154 M:112 C: (48) String data converted to lower case,

M:112 Repeated in SeqNo=3

L:208 M:112 C: (48) String data converted to lower case,

M:112 Repeated in SeqNo=4